

2015-02-23

The Value of Authority Control in a Next Generation OPAC

Ahronheim, Judith; Case, Beau; DeGeorge, Deborah; Sears, JoAnn

<https://hdl.handle.net/2027.42/110677>

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Title

The Value of Authority Control in a Next Generation OPAC

Authors

Judith Ahronheim, Beau David Case, Deborah S. DeGeorge, and JoAnn Sears.

Abstract

This white paper, based upon the final report of the University of Michigan Library Authority Control Working Group, submitted March 30, 2011, demonstrates that, within an environment of a single, Google-like search box, results are far more accurate with controlled headings.

Introduction

The developing practice of large quantities of heterogeneous bibliographic data into database indexes accessed through a single interface raises questions about what place controlled headings might have in this new environment. At the University of Michigan Library, the following conditions currently exist:

- VuFind (1.0), the Library's catalog interface, does not index cross references. This limits the value of authority control since users are not referred from name and subject variants and such variants would not be retrieved unless they happened to appear in an indexed free-text field;
- New records are outsourced to an authority control vendor to process them and update headings to conform to authorized forms. Existing records are updated when the national authorized forms are changed;
- The library is considering including the contents of its catalog in a single search Summon interface alongside indexed articles;
- The library spends money, time and labor on maintaining authority control, which may not be of value to users in modern search interfaces.

Does a case still exist for providing enriched access via the management of authorized forms of names, topics, and geographic areas and their variants in the catalog? To provide some data to consider in answering the question of authority control value, the authors conducted a variety of data-gathering investigations:

- Literature review;
- Practices at other institutions;
- Log data and statistics from the catalog;
- Poll of Library staff;
- Examination of one-week random sample of user searches.

Assigning value to library efforts is challenging; some libraries have attempted to do this by measuring the impact, or return, on their budget. Historically, public libraries have utilized this more than other types; the American Library Association (ALA) maintains a listing with several examples. More recently, academic libraries also have started to employ the return on investment (ROI) model and the social return on investment (SROI) model. Luther and Tenopir provide some recent work in this area, analyzing the overall return on research dollars at a university and the impact of electronic journals on a campus community.¹ Related to these models is a contingent valuation (CV) analysis, which is reviewed for

library settings by Missingham. Our work here includes a poll of library staff that aligns with the CV methodology.

Literature review

A literature review was conducted to determine current research on authority control. The document, *Final Report of the Task Force on Cost/Value Assessment of Bibliographic Control*, suggests that serious investigation into the costs and benefits of modern authority control is in the early stages, particularly as the “value” of authority control is difficult to quantify.² The report further suggests that the ALA Heads of Technical Services of Large Research Libraries (to whom the report was submitted), in conjunction with vendors and bibliographic utilities, should take up said research.

Works by Salo, Habibzadeh, Rafferty, Blake, and Harper suggest that managers of databases other than library catalogs are beginning to see the need for normalization and syndetic structures of some kind. Salo discusses the retrieval problems caused by lack of controlled names in institutional repositories, noting that the numerous variations on authors’ names prevent the user from locating all articles by any one author.³ Habibzadeh, writing about scientific literature, notes that the lack of unique author identification affects the impact factor in scientific publications; this is especially problematic for researchers for whom this impact factor will affect career promotion, grants and other funding, and recruitment.⁴ Rafferty discusses the issues with “author-indexed databases,” those with author-generated, uncontrolled metadata: most prevalent are overly-broad terms, ambiguous terms, and the inclusion of synonyms or misspellings.⁵ Blake writes that controlled headings for organizational names are necessary in an electronic resource management system (ERMS), particularly when names are imported into the ERMS from a variety of sources and formats. Without controlled name forms, difficulty arises in license mapping and report generation.⁶ Harper discusses authority control in terms of the semantic web, noting that data from the Virtual International Authority File (VIAF) could be linked to web-based resources such as Wikipedia pages, thus improving precision. The authors also mention non-library-based vocabularies for names, such as the Friend of a Friend (FOAF) project, which can be used for aggregating information.⁷

From a more catalog-focused perspective, Wilkes’s 1995 study compares the results of subject searching in two online catalog systems, one system with authority control, the other without:⁸

Transaction logs from Library A (no authority control) were analyzed to identify searching patterns of users; 885 searches were attempted, 351 (39.7%) by subject. One hundred forty-two (40.6%) of these subject searches were unsuccessful. Identical searches were performed in a comparable library that had authority control, Library B. Terms identified in “see” references at Library B were searched in Library A. One hundred five (73.9%) of the searches that appeared to fail would have retrieved at least one, and usually many, records if a link had been provided between the term chosen by the user and the term used by the system.

A fairly recent survey of authority control practices was published in June 2010, and was used by the authors of this paper to identify the extent of authority control conducted at our institutional equivalents.⁹

Practices at Other Institutions

The following charts were extracted from a report of a survey published in 2010 in *Library Resources & Technical Services*.¹⁰ The authors conducted a randomized survey and collected and organized data to an extent that the authors of this paper would not have been able to achieve by polling colleagues:

Table 2. Cataloging and Name Authority Practice by Institution Type and Collection Size

	Graduate % (n = 39)*	4-Year % (n = 35)*	2-Year % (n = 24)*	Largest % (n = 30)*	Middle % (n = 30)*	Smallest % (n = 30)*
Cataloging						
Almost all done in-house	58	71	88	43	90	80
Most done in-house, some outsourced	30	9	0	40	0	3
Proportion in-house/outsourced about even	8	11	0	10	3	7
Most outsourced, some done in-house	3	3	8	3	7	3
Almost all outsourced	3	6	4	3	0	7
Name Authority Work Done						
Always	31	29	29	30	43	23
Usually	33	26	38	40	27	30
Sometimes	28	29	21	27	20	30
Never	8	17	13	3	10	17
NACO-Authorized	35	25	0	41	16	5

Table 3. Types of Name Authority Work Done by Institution Type and Collection Size

	Graduate % (n = 39)*	4-Year % (n = 35)*	2-Year % (n = 24)*	Largest % (n = 30)*	Middle % (n = 30)*	Smallest % (n = 30)*
Types of Names Controlled						
Personal	100	93	95	100	100	92
Corporate	83	73	80	93	89	58
Geographic	78	7	80	90	85	54
Series	78	73	55	90	67	58
Conference	75	63	55	83	82	42
Uniform titles	69	47	50	76	56	38
Types of Materials Controlled						
Monographs	100	97	95	100	100	92
Other nonbook materials	75	72	62	90	74	48
Serials	69	48	43	72	56	40
Electronic materials	58	55	29	69	63	20

Table 4. Name Authority Processes by Institution Type and Collection Size

When Name Authority Work is Done	Graduate % (n = 39)*	4-Year % (n = 35)*	2-Year % (n = 24)*	Largest % (n = 30)*	Middle % (n = 30)*	Smallest % (n = 30)*
Precataloging	3	0	5	3	0	4
During cataloging	57	50	68	31	64	77
Postcataloging	41	50	27	66	36	19
Always/Usually Check Verification Sources						
Original cataloging	82	79	75	87	80	69
Copy cataloging	46	47	54	40	70	45
On records received from vendors	29	43	57	35	50	42
Verification Sources Used						
OCLC	84	79	73	90	86	69
LC online authority file	46	52	55	41	55	58
Local authority file	43	36	36	31	38	52
Regional cataloging system	0	12	9	3	0	12
Other libraries' databases	3	9	9	3	3	15
Vendor files	11	0	14	10	14	0
Used for Updating Existing Headings						
Staff and patron suggestions	56	43	55	48	50	59
Heading changes listed by LC	50	46	60	55	54	55
Reports generated by online catalog	47	46	55	62	46	41
Periodic file maintenance by a vendor	34	39	30	38	38	18
Vendor notifications of heading changes	31	11	20	35	21	5
Authority Control Module (ACM)						
ACM actively integrated with bibliographic file	53	59	50	60	57	47
ACM not actively integrated	26	18	21	30	23	17
Library catalog has ACM, but not using it	13	9	4	10	3	10
Library catalog does not have ACM	8	15	25	0	17	27
Library has no computerized catalog	0	0	0	0	0	0

*Note: This is the overall n; however there was slight variation per item for n, as not every respondent answered every question.

Poll of Library staff

A variety of individuals and groups throughout the library system were surveyed to assess opinions of cross reference use and value. Data was collected using paper surveys (see Appendices B and C) and group discussion. The majority of those surveyed were public services librarians, with approximately 20 percent of the responses coming from technical services staff. See Appendix B for the specific questions asked during these interviews. Responses were coded individually, but also summarized by broad subject area to look for patterns across units.

The majority (75 percent) of survey respondents indicated that they regularly used cross references in the university's online catalog. That same majority of people affirmed that they had even made use of cross references recently (i.e., in the last month). The number jumped to 89 percent when respondents were asked if they had clicked on links provided in the catalog to get related records, providing some evidence that people use title/series/author authority control even if they do not identify these as authorized forms. The group that did not exhibit majority use of cross references was the health sciences library staff. In the interviews of library staff, examples of cross references were provided (See Appendix C). Staff were asked about the frequency of use of these types of authority control work. An overwhelming 80percent of staff across all areas marked "often" or "sometimes," while the remaining 20 percent marked "rarely," "never," or "n/a". No subject/work area deviated from having a majority in the first two categories.

Finally, staff were asked about the perceived monetary value of cross references. The majority of staff marked \$0.10 or \$1.00 per cross reference, as demonstrated by this table:

Value per x –ref:	\$0.01	\$0.10	\$1.00	\$10.00	n/a
Number of staff:	4	20	21	16	15

Additional sample comments from staff during these interviews are included below; these mainly reflect a response to the question: *How would a lack of authority control affect your work?*

- “Most of my day-to-day work would take several times as long—I’d have to do a lot more research to know that what I was looking for matched other instances of that person/corporation/subject.” (Technical Services)
- “Would make searching in [the catalog] a much more tedious process. In some cases, it would lead to duplicating items in the collections; providing inaccurate information to patrons—referring more to ILL when items may already be in the collection.” (Humanities)
- “It would make it much harder; it would probably mean some reference questions might not be answerable.” (Undergraduate Library)
- “Lack of authority control would make it very difficult to track works by an author who publishes in multiple languages, who Romanize their names differently from library standards, differentiate authors of similar names (a very common problem in Chinese and Korean authors). For subject headings, it would eliminate one major tool I direct students to use when searching.” (International Studies)
- “I would not be able to answer reference questions as quickly. It would also be harder for me to do my own research and to help others who are consulting me about their research.” (Social Sciences)
- “It would make our catalog less searchable and browsable; it would probably lead to confusion between authors with similar names.” (Arts)
- “I would cry a thousand tears.” (Engineering)
- “In medical reference, not much, b/c we spend more time w/ journal lit and don't often search [the catalog] for subjects. At another library, I use subjects extensively in general ref.” (Health Sciences)
- “Author lack of authority control would be particularly difficult & impact my work in a negative way.” (Science)
- “My work is already impacted by a lack of authority control--geospatial metadata doesn't even exist much of the time, much less have authority control.” (Social Sciences)
- “Without authority control, I would be unable to perform my job. For starters, assigning call numbers for scores requires searching for all other editions and works related to the item in hand. Especially when dealing with uniform titles (as is most often the case with musical works), a lack of authority control would make this impossible.” (Technical Services)
- “I already have found that lack of authority control in series titles is a bit of a pain, but it’s tolerable. Authority control for authors/contributors is essential, as are uniform titles.” (Arts)
- “Hard to say. All of the authority control work has its purpose; the more we do, the more accurate reference work and patron searching can be done. The lack of authority work with serials is already creating confusion.” (Humanities)
- “I think we could maybe do without title authority control (including uniform titles). Series authority control has already been more or less done away with and I agree with that. Need personal and corporate name authority control.” (Special Collections)

Examination of user searches

Three types of user searching behavior were observed to determine if authority control aided in retrieval.

The authors first retrieved raw number search and scan (i.e., browse) requests for the period of July 2008 through December 2008 from the local implementation of the ExLibris catalog.

The results are as follows:

Search ISBN	889
Search ISSN	7,729
Search Words	46,501
Search Titles	489
Search Title Words	4,788
Search Author Words	891
Search Subject Words	622
Search OCLC Number	305
Search MESH Words	5
Search Journal Title	323
Search System Number	30
Search Utility Number	2,941
Search Serial Title Words	1,301
Search Journal Title Words	8
Search MDP Number	233
Scan Author	15,568
Scan Chinese Author	22
Scan Japanese Author	21
Scan Korean Author	19
Scan Call Number	3,666
Scan Other Call Number	11
Scan Sudoc Number	9
Scan ISBN	2,176
Scan ISSN	3,787
Scan MESH	123
Scan Series Title	772
Scan Journal Title	10,040
Scan LCSH	4,751
Scan Title	32,129
Scan Chinese Title	57
Scan Japanese Title	8
Total	140,214

Although only raw data were retrieved rather than exact user searches, the data are rich and interesting. A high percentage of all Mirlyn Classic use from that period (searches and browses combined) was for browses, a type of searching utilizing authority control: 73,129 (52.18 percent) of the 140,214 uses were browses. Within the browses, searches alone, these catalog searches accounted for a significant portion of the total use: Authors 15,586/140,214 (11.10 percent), Titles 32,129/140,214 (22.91 percent), Journal Titles 10,040/140,214 (7.16 percent).

Searches from both the library's website and Mirlyn were retrieved from the period October 24, 2008 through October 31, 2008. A random sample of 955 searches from these datasets was created, divided into three groups (313; 320; 322¹¹), and searched in Mirlyn Classic to determine if authority control aided in result retrieval. The Mirlyn Classic searches were then searched again in the Browse tab: each of the random searches was searched as "Title begins with," as "Author," and as "Subject begins with." If an authorized form result or cross-reference was retrieved in any of the browses, the random search sample was counted as a positive hit; if no result was retrieved in any of the three browse categories, then the random search sample was counted as not being a hit.

The browse results were as follows:

	Number	Positive hits	Percentage of positive hits from sample
Sample 1	313	176	0.56
Sample 2	320	203	0.63
Sample 3	322	112	0.35
totals	955	491	0.51

The striking result from this random search data is that more than half of all the user searches retrieved results when browsed against a library catalog utilizing the syndetic cross-reference structure common within authority control.

Third, the set of 313 randomly selected searches was researched in VuFind and compared with the results from Mirlyn Classic. Forty-one (13.10 percent) of these comparisons produced results in Classic that were either not displayed or so mixed as to be irretrievable in VuFind.¹²

Conclusion

Our investigations suggest that there is still a place for formal authority control in library catalogs. While many user needs may be met with simple uncontrolled results, there is a significant percentage where the value and precision added by authority control, at most, makes possible and, at least, enhances useful results when user-conducted searches fail. This is indeed a value that the library adds to the retrieval and relevance of search results.

Appendix A: Survey

- Do you use cross references?
- Have you used one in the last month?
- In the last month, have you clicked on a subject, author title, or series in a display in order to get to related records?
- Based upon the examples provided, how often do you encounter similar situations:
Often / Sometimes / Rarely / Never?
- In your opinion, what is the monetary value of an individual cross reference?
\$.01 / \$.10 / \$1.00 / \$10.00?
- How would a lack of authority control affect your work?
- If we limited the amount of authority control, what could you do without? What would be a show-stopper?

Appendix B: Examples

A1. *Clemens, Samuel* Langhorne (1835 – 1910)

A2. DMCA

A3. Madonna

A4. Mercury

B1. A Golden Book

B2. Bulletin

B3. Complete Works

B4. Garland Library

C1. Afro-American

C2. Indians

C3. Labored Breathing

C4. Metal Locators

Appendix C: Sample of user searches

proquest
learning to think
new yorker
marketing
cat
stereotypes in media
unity
geometry of surfaces
moments in blue
"peace corp"
michael reisch
swine flu
gallant tom
the craft of bureaucratic neutrality
robert g. mccloskey
linkage of early-onset familial breast cancer to chromosome
17q21
helen longino
comm and mass media complete
z21.11
1001 solved problems
pavilion church
race, ethnicity, and disability: veterans and benefits in post-civil
war america
los peces de amargura
the jews of eastern europe
writer's resource
picasso
el arte de la resurreccion
a cameo from cranford
acs
into the woods
color of night
puerto rican separatists
the crazies
carbonated soft drinks:
the giving tree
critical race theory
like water for chocolate
sleepy hollow
drupal

increasing access to college: extending possibilities for all students

puerto rican revolutionary nationalism

quaternary climate

zizek

tauber, warren bryan

fems microbiol lett

1046-4883

subtractive schooling

george starr lasher

william tierney

Appendix D: Examples of differing results in Mirlyn Classic and VuFind

Rösberg (Bornheim, Rhein-Sieg-Kreis, Germany) History	only located via authority-controlled heading
same sex marriage	176 authority-controlled headings; VuFind does not display ABC poll data under Gay marriage.
rudolph steiner	name misspelled. 17 hits in VuFind; 100+ in Classic by author.
Rossini	1593 hits found (authority-controlled facets in VuFind show 2 identities; actually there are 48 distinct Rossini identities in Classic)
Rockwell, D	370 hits in VuFind; only DM Rockwell appears on first page. Daisy, David, David C. and Donald (from Classic) are not apparent on first page or in facets
Robert Graves	997 entries in VuFind; 16 distinct identities found in Classic; only two in VuFind facets
que	29,303 hits in VuFind results; Que series title and 25 author identities in Classic do not appear in first screen or facets in Vufind.
proceedings on the 8 International Congress of Greek and Latin Epigraphy	found because English title is found in authority controlled heading; actual title is in Greek
Peter James	6119 hits in VuFind results; 14 distinct identities found in Classic, only one appears in VuFind first page, none in facets
ohanyan	7 identities in Classic, only 5 show up in VuFind
national geographic	journal does not appear in first page of 3853 VuFind hits nor in facets
Millennials	1 title in VuFind; X-ref to Generation Y in Classic shows 6 results
mathematics	123,264 hits in VuFind; Classic subject headings provide better focus—also easier to find journal title
john pawson	3 John Pawson identities in Classic, only 2 show up in VF facets
jarva	3 identities in Classic, only two retrieved in Vufind
J. Chem. Ed.	VuFind : 120,644 results, J Chem Education is not on first VuFind page; has heading in Classic
Wininger	4 Winingers distinguished
the medical journal of austrialia	misspelling: australia appears in Classic browse
television	subject focus provided by authority control
Teen Pregnancy	additional titles and focus provided by authority control

References

- ¹ American Library Association. "Articles and Studies Related to Library Value (Return on Investment)," include an accessed date before the URL, www.ala.org/research/librarystats/roi.
- ² ALA Heads of Technical Services in Large Research Libraries Interest Group Task Force on Cost/Value Assessment of Bibliographic Control. *Final Report of the Task Force on Cost/Value Assessment of Bibliographic Control*, June 18, 2010.
http://connect.ala.org/files/7981/costvaluetaskforcereport2010_06_18_pdf_77542.pdf
- ³ Dorothea Salo, "Name Authority Control in Institutional Repositories," *Cataloging & Classification Quarterly* 47, no. 3-4 (2009), 249-261.
- ⁴ Farrokh Habibzadeh, "The Problem of 'Who,'" *The International Information & Library Review* 41, no. 2 (2009), 61-62.
- ⁵ Pauline Rafferty, "Flickr and democratic indexing : dialogic approaches to indexing," *ASLIB Proceedings : New Information Perspectives*, v. 59, no. 4-5 (2007), .4.
- ⁶ Kristen Blake, "Creating Organization Name Authority within an Electronic Resources Management System," *Library Resources & Technical Services* 53, no. 2 (2009), 94-107.
- ⁷ Corey A. Harper and Barbara Tillett, "Library of Congress Controlled Vocabularies and their Application to the Semantic Web," *Cataloging & Classification Quarterly* 43, no. 3-4 (2009), 47-68.
- ⁸ Adeline Wilkes, "Subject searching in Two Online Catalogs: Authority Control vs. Non-Authority Control," *Cataloging & Classification Quarterly* 20, no. 4, (1995), 57-79.
- ⁹ Susan K. Burke and Jay Shorten, "Name Authority Work Today: A Comparison of Types of Academic Libraries," *Library Resources & Technical Services* 54, no. 1 (2010), 4-20.
- ¹⁰ Ibid.
- ¹¹ See Appendix C for examples of searches.
- ¹² See Appendix C for examples of searches.